

Package ‘CFM’

November 14, 2025

Type Package

Title Analyzing Censored Factor Models

Version 0.7.0

Description Provides generation and estimation of censored factor models for high-dimensional data with censored errors (normal, t, logistic). Includes Sparse Orthogonal Principal Components (SOPC), and evaluation metrics. Based on Guo G. (2023) <[doi:10.1007/s00180-022-01270-z](https://doi.org/10.1007/s00180-022-01270-z)>.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

Depends R (>= 3.5.0)

Imports stats, MASS, psych, matrixcalc, crch, Matrix, pracma,
RSpectra, kernlab, mvtnorm, robustbase, truncnorm

Suggests testthat (>= 3.0.0), ggplot2

NeedsCompilation no

RoxygenNote 7.3.3

Author Guangbao Guo [aut, cre],
Tong Meng [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Repository CRAN

Date/Publication 2025-11-14 00:00:08 UTC

Contents

censored_factor_model	2
censored_factor_models	3
censored_kernel_factor_model	4
CFM	4
FanPC.CFM	5
incremental_censored_factor_model	7
PC2.CFM	8

PPC2.CFM	9
weighted_censored_factor_model	10
Index	12

censored_factor_model	<i>Robust Censored Factor Model</i>
-----------------------	-------------------------------------

Description

Implementation of robust censored factor model using robust PCA initialization for handling left-censored data.

Usage

```
censored_factor_model(  
  X,  
  m,  
  max_iter = 100,  
  tol = 1e-04,  
  nugget = 1e-06,  
  alpha = 0.75  
)
```

Arguments

X	Data matrix (n x p)
m	Number of factors
max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)
alpha	Robustness parameter for MCD (default: 0.75)

Value

A list containing model results

`censored_factor_models`*Basic censored-factor data simulator*

Description

Generates multivariate data that follow a latent factor structure with censored errors (Normal, Student-t or Logistic).

Usage

```
censored_factor_models(  
  n,  
  p,  
  m,  
  distribution = c("normal", "t", "logistic"),  
  df = NULL,  
  seed = NULL  
)
```

Arguments

<code>n</code>	Sample size (> 0).
<code>p</code>	Number of observed variables (> 0).
<code>m</code>	Number of latent factors ($< p$).
<code>distribution</code>	Error distribution: "normal" (default), "t", "logistic".
<code>df</code>	Degrees of freedom when distribution = "t".
<code>seed</code>	Optional random seed.

Value

A list with components:

<code>data</code>	numeric $n \times p$ matrix of observations
<code>loadings</code>	$p \times m$ factor loadings matrix
<code>uniqueness</code>	$p \times p$ diagonal uniqueness matrix
<code>KMO</code>	KMO measure of sampling adequacy
<code>Bartlett_p</code>	p-value of Bartlett's test
<code>distribution</code>	error distribution used
<code>seed</code>	random seed

Examples

```
set.seed(2025)  
obj <- censored_factor_models(200, 6, 2)  
psych::KMO(obj$data)
```

censored_kernel_factor_model
<i>Kernel Censored Factor Model</i>

Description

Implementation of kernel-based censored factor model using kernel PCA initialization for nonlinear factor analysis with censored data.

Usage

```
censored_kernel_factor_model(  
  X,  
  m,  
  kernel_type = "rbf",  
  gamma = NULL,  
  max_iter = 100,  
  tol = 1e-04,  
  nugget = 1e-06  
)
```

Arguments

X	Data matrix (n x p)
m	Number of factors
kernel_type	Kernel type: "rbf" or "linear" (default: "rbf")
gamma	Gamma parameter for RBF kernel (default: 1/p)
max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)

Value

A list containing model results

CFM	<i>Censored Factor Models Data Generation</i>
-----	---

Description

Generate multivariate data that follow a latent factor structure with censoring errors drawn from Normal, Student-t or Logistic distributions. Convenience wrapper around [rcnorm](#), [rct](#), and [rclogis](#).

Usage

```
CFM(n, p, m, cens.dist = c("normal", "t", "logistic"), df = 5, seed = NULL)
```

Arguments

n	sample size ($n \times 1$ observations).
p	number of manifest variables.
m	number of latent factors.
cens.dist	censoring error distribution: "normal", "t", or "logistic".
df	degrees of freedom when cens.dist = "t".
seed	optional random seed for reproducibility.

Value

A named list with components:

data	numeric $n \times p$ matrix of observations.
F	factor scores matrix ($n \times m$).
A	factor loadings matrix ($p \times m$).
D	unique variances diagonal matrix ($p \times p$).

Examples

```
set.seed(2025)
# Normal censoring
obj <- CFM(n = 200, p = 10, m = 3, cens.dist = "normal")
head(obj$data)

# t-censoring with 6 d.f.
obj <- CFM(n = 300, p = 12, m = 4, cens.dist = "t", df = 6)
psych::KMO(obj$data)
```

Description

Censored Factor Analysis via Principal Component (FanPC, pure R)

Usage

```

FanPC.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)

```

Arguments

<code>data</code>	Numeric matrix or data frame of dimension $n \times p$.
<code>m</code>	Number of factors ($< p$).
<code>A</code>	Optional true loading matrix, used only for error calculation.
<code>D</code>	Optional true unique-variance diagonal matrix, used only for error calculation.
<code>p</code>	Number of variables (deprecated; detected automatically).
<code>cens.dist</code>	Error distribution, reserved for future use.
<code>df</code>	Degrees of freedom, reserved for future use.
<code>cens.method</code>	Censoring handling method; currently only "winsorise" is implemented. Defaults to "winsorise".
<code>cens_prop</code>	Winsorisation proportion, default 0.01.
<code>surv.obj</code>	Reserved for future use.
<code>ctrl</code>	Reserved for future use.
<code>verbose</code>	Reserved for future use.

Value

AF Estimated loading matrix, $p \times m$.

DF Estimated unique-variance diagonal matrix, $p \times p$.

MSESigmaA Mean squared error of loadings (if A is provided).

MSESigmaD Mean squared error of unique variances (if D is provided).

LSigmaA Relative error of loadings (if A is provided).

LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 10, m = 2, cens.dist = "normal")
res <- FanPC.CFM(obj$data, m = 2, A = obj$A, D = obj$D, cens.method = "winsorise")
print(res$MSESigmaA)
```

incremental_censored_factor_model

Incremental Censored Factor Model

Description

Implementation of incremental censored factor model for streaming data with left-censored observations. Processes data in batches.

Usage

```
incremental_censored_factor_model(
  batch_data,
  m,
  max_iter = 100,
  tol = 1e-04,
  nugget = 1e-06
)
```

Arguments

batch_data	List of data matrices (batches)
m	Number of factors
max_iter	Maximum number of ECM iterations (default: 100)
tol	Convergence tolerance (default: 1e-4)
nugget	Numerical stability term (default: 1e-6)

Value

A list containing model results

PC2.CFM

*PC2 for censored factor models (Top-2 principal components, pure R)***Description**

PC2 for censored factor models (Top-2 principal components, pure R)

Usage

```
PC2.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)
```

Arguments

<code>data</code>	Numeric matrix or data frame of dimension $n \times p$.
<code>m</code>	Number of factors ($< p$).
<code>A</code>	Optional true loading matrix, used only for error calculation.
<code>D</code>	Optional true unique-variance diagonal matrix, used only for error calculation.
<code>p</code>	Number of variables (deprecated; detected automatically).
<code>cens.dist</code>	Error distribution, reserved for future use.
<code>df</code>	Degrees of freedom, reserved for future use.
<code>cens.method</code>	Censoring handling method; currently only "winsorise" is implemented. Defaults to "winsorise".
<code>cens_prop</code>	Winsorisation proportion, default 0.01.
<code>surv.obj</code>	Reserved for future use.
<code>ctrl</code>	Reserved for future use.
<code>verbose</code>	Reserved for future use.

Value

AF Estimated loading matrix, $p \times 2$.

DF Estimated unique-variance diagonal matrix, $p \times p$.

MSESigmaA Mean squared error of loadings (if A is provided).

MSESigmaD Mean squared error of unique variances (if D is provided).

LSigmaA Relative error of loadings (if A is provided).

LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 12, m = 2, cens.dist = "normal")
res <- PC2.CFM(obj$data, A = obj$A, D = obj$D)
print(res$MSESigmaA)
```

PPC2.CFM

PPC2 for censored factor models (Top-2 principal components, pure R)

Description

PPC2 for censored factor models (Top-2 principal components, pure R)

Usage

```
PPC2.CFM(
  data,
  m,
  A = NULL,
  D = NULL,
  p = NULL,
  cens.dist = c("normal", "t", "logistic"),
  df = NULL,
  cens.method = c("winsorise", "em"),
  cens_prop = 0.01,
  surv.obj = NULL,
  ctrl = NULL,
  verbose = NULL
)
```

Arguments

<code>data</code>	Numeric matrix or data frame of dimension $n \times p$.
<code>m</code>	Number of factors ($< p$).
<code>A</code>	Optional true loading matrix, used only for error calculation.
<code>D</code>	Optional true unique-variance diagonal matrix, used only for error calculation.
<code>p</code>	Number of variables (deprecated; detected automatically).
<code>cens.dist</code>	Error distribution, reserved for future use.
<code>df</code>	Degrees of freedom, reserved for future use.
<code>cens.method</code>	Censoring handling method; currently only "winsorise" is implemented. Defaults to "winsorise".
<code>cens_prop</code>	Winsorisation proportion, default 0.01.
<code>surv.obj</code>	Reserved for future use.
<code>ctrl</code>	Reserved for future use.
<code>verbose</code>	Reserved for future use.

Value

AF Estimated loading matrix, $p \times 2$.
DF Estimated unique-variance diagonal matrix, $p \times p$.
MSESigmaA Mean squared error of loadings (if A is provided).
MSESigmaD Mean squared error of unique variances (if D is provided).
LSigmaA Relative error of loadings (if A is provided).
LSigmaD Relative error of unique variances (if D is provided).

Examples

```
library(CFM)
obj <- CFM(n = 500, p = 12, m = 2, cens.dist = "normal")
res <- PPC2.CFM(obj$data, A = obj$A, D = obj$D, cens.method = "winsorise")
print(res$MSESigmaA)
```

weighted_censored_factor_model

Weighted Censored Factor Model

Description

Implementation of weighted censored factor model with weighted PCA initialization for handling left-censored data with observation weights.

Usage

```
weighted_censored_factor_model(  
  X,  
  m,  
  weights = NULL,  
  max_iter = 100,  
  tol = 1e-04,  
  nugget = 1e-06  
)
```

Arguments

<code>X</code>	Data matrix (n x p)
<code>m</code>	Number of factors
<code>weights</code>	Observation weights vector of length n (optional)
<code>max_iter</code>	Maximum number of ECM iterations (default: 100)
<code>tol</code>	Convergence tolerance (default: 1e-4)
<code>nugget</code>	Numerical stability term (default: 1e-6)

Value

A list containing model results

Index

censored_factor_model, [2](#)
censored_factor_models, [3](#)
censored_kernel_factor_model, [4](#)
CFM, [4](#)

FanPC.CFM, [5](#)

incremental_censored_factor_model, [7](#)

PC2.CFM, [8](#)
PPC2.CFM, [9](#)

rclogis, [4](#)
rcnorm, [4](#)
rct, [4](#)

weighted_censored_factor_model, [10](#)